

# **Report 1 (R1)**

## **An Overview of the 2002 Commodity Origin-Destination Database: Methodology and Data**

The 2002 Commodity Origin-Destination Database is comprised of three four-dimensional matrices (for tons, ton miles, and value) in which the four dimensions are origin, destination, commodity, and mode. Origins and destinations consist of 114 regions as defined and used in the 2002 Commodity Flow Survey (CFS) plus 17 international gateways and 7 international regions. Commodities are defined at the 2-digit SCTG (Standard Classification of Transported Goods) level. Modes are defined as in the 2002 CFS – i.e. 11 separate modes, multimodal combinations, and unknown modes—but reported for only 7 aggregated modes in FAF. The 2002 CFS serves as the foundation of the 2002 Commodity Origin-Destination Database. However, the CFS has several major commodity gaps, referred to as out-of-scope commodities. In addition, the CFS undercounts some categories of trade and movements of freight, for example, in-transit movements, petroleum products, and exports. These CFS out-of-scope commodities and undercounts are addressed by ORNL and MacroSys in a series of special reports.

### **Data Overview**

The 2002 Commodity Origin-Destination Database is derived from three categories of data: CFS Within-Scope Data, Auxiliary Data, and CFS Out-of-Scope Data.

#### **CFS Within-Scope Data**

CFS Within-Scope Data at the state level comes from CFS Table 17. This information is only available on CD from the Bureau of Transportation Statistics ([answers@bts.gov](mailto:answers@bts.gov) or 800-853-1351).

STATE Table 17. Shipment Characteristics by Destination State, Two-Digit Commodity and Mode of Transportation for State of Origin: 2002

At this level of detail, the majority of cells are empty or suppressed for one or more reasons, including disclosure rules and suppression due a lack of statistical significance.

#### **Auxiliary Data**

The very spare nature of CFS Table 17 calls for additional information from other sources about missing cells. This additional information, termed Auxiliary Data, comes from various sources and is used in the Log-linear Model to estimate effects at the 2, 3,

and 4 dimensional levels of the 2002 Commodity Origin-Destination Database. Please see Report No. 3 for details.

### **Census Bureau “True-Zero” CFS Cells:**

FHWA and ORNL issued a special request to the U.S. Census Bureau to identify those cells within CFS Table 17 for which there were no observations per their 2002 survey. With the strong cooperation of Census, this request was filled.

Per our methodology, FHWA and ORNL assumed that any cell within Table 17 for which the 2002 CFS survey had no responses is declared a “true zero.” The combination of Table 17 with this “True Zero” version of Table 17 allows many of the cells in the 2002 Commodity Origin-Destination Database to be identified. This is of value during the Log-linear and Iterative Proportional Fitting steps by which gaps in Table 17 (disaggregated from the state level to the FAF level) are filled. Please see Report No. 3 for details.

### **Database: Carload Waybill Sample**

The Surface Transportation Board's (STB) Carload Waybill Sample is a stratified sample of carload waybills for terminated shipments by U.S. railroad carriers. The Association of American Railroads (AAR) collects Waybill data annually for STB from railroads that have moved at least 4,500 carloads each year for each of the previous three years, or which move 5% or more of any State's total rail traffic. Sample stratification is based on the number of railcars a railroad moves and on the number of carloads in a movement. Waybills reporting large number of carloads, such as unit train movements involving more than 100 carloads, have a higher probability of selection than smaller movements.

AAR generates both a Public Use waybill sample, and a more detailed dataset for the same sample that is restricted to internal government use. The Public Use File provides estimates of annual origin-to-destination tonnages and revenues received by specific railroads at the State-to-State and BEA (Bureau of Economic Analysis) region-to-region level. Commodities are reported at the 5-digit level using STCC (Standard Transportation Commodity Codes). The restricted dataset incorporates added geographic detail for both O-D identification and railway routing. This more detailed dataset can be used, for example, to assign annual O-D rail flows to specific FAF Regions. It also can be used to improve the selection of specific routings for the purposes of rail traffic assignment (and subsequent rail ton-mileage estimates). Expansion factors are provided for both datasets that allow users to expand the sample data to national totals. While the sample covers all commodities carried by in-scope U.S freight railroads, it does not capture export shipments carried on Canadian railroads operating inside the United States.

For FAF use, STCC commodity codes had to be converted to SCTG codes. This was done by assigning each 5-digit STCC code to a specific 2-digit SCTG commodity class. See Appendix #3 for crosswalk details.

Information on these data are found at:

Railinc (2002) User Guide for the 2002 Surface Transportation Board Carload Waybill Sample. Association of American Railroads, July 31, 2002.

[http://www.stb.dot.gov/stb/industry/econ\\_waybill.html](http://www.stb.dot.gov/stb/industry/econ_waybill.html)

#### **Database: Domestic Waterborne Commerce of the United States**

The U.S. Army Corps of Engineers (USACE) provided data on U.S. waterborne commerce, including the transport of goods by inland barge and ship over the nation's navigable rivers, across the Great Lakes, and within the U.S. Intra-Coastal Waterway. Domestic O-D movements are created by USACE from its Vessel Operating Reports, as well as from its Lock Performance Monitoring System (LPMS) database. Data are in theory reported by all vessels and provide estimates of annual tons moved by 5-digit commodity code for all commodities transported on U.S. waterways, on a dock-to-dock basis. These data were aggregated geographically and used to supply the FAF with State-to-State as well as FAF Region-to-Region annual commodity tonnage totals. For this purpose the data was converted from 4-digit Waterborne Commerce (WCUS) commodity codes to 2-digit SCTG commodity classes. Please see Appendix #4 for crosswalk details.

Information on these data are found at:

<http://www.iwr.usace.army.mil/ndc/wcsc/wcsc.htm>

#### **Database: International Waterborne Commerce of the United States**

USACE, International Waterborne Transportation Statistics Program, provided international waterborne commerce data, which is based on information supplied to USACE by the U.S. Census Bureau. These data cover vessels engaged in U.S. foreign trade and transportation, including cargo data by type of service, by U.S. and foreign port, by country of origin/destination, commodity, value and tonnage, for both bulk and containerized cargo. Data are provided in accordance with the Harmonized Schedule (HS) of reporting. A conversion from HS to SCTG commodity classes was carried out for FAF use, as was an assignment of foreign counties to the 7 FAF Foreign Regions. Please see Appendix #5 for crosswalk details.

Information on these data are found at:

<http://www.iwr.usace.army.mil/ndc/usforeign/index.htm>

The import and export data are found at:

<http://www.iwr.usace.army.mil/ndc/db/foreign/data/>

### **Database: Transborder Surface Freight**

The Bureau of Transportation Statistics provided the Transborder Surface Freight database, which includes data from the U.S. Customs Service, via the U.S. Census Bureau. These data provide the FAF with the dollar value of both imports and exports at U.S.–Canadian and U.S. Mexican land border crossings, as well as the tonnage of imports. These data are broken down according to truck, rail, pipeline, mail and other moves by the 2-digit harmonized Schedule (HS) of commodity classes. Geographically this O-D data is broken down by U.S. State, Canadian Province, and Mexican State of origin and destination (Mexican state of origin for U.S. imports is not reported). The BTS public domain database also identifies total annual mode specific movements through each U.S. port of entry or exit by U.S. state of origin or destination. Imports valued at less than \$1,250 and exports valued at less than \$2,500 are not included in these data, nor are transshipments.

Information on these data are found at:

<http://www.bts.gov/transborder/>

### **Database: U.S. Air Freight Movements**

The volume (payload weight) and O-D pattern of domestic and international revenue-generating air freight within the United States are available from the Office of Airline Information (OAI), Bureau of Transportation Statistics. The data used in FAF is taken from Form 41 Air Carrier, the T-100 Market Data. These data report the annual payload tons of mail as well as freight flown between each pair of U.S. airports over the course of a year. No commodity disaggregation is available, nor is any data on the value of the freight involved. The database identifies the State of originating U.S. airport and State of destination U.S. airport for these cargo movements.

Information on these data can be found by selecting ‘Aviation’ and ‘Air Carrier Statistics (Form 41 Traffic)’ once at the following website:

<http://www.transtats.bts.gov/>

Both combined and separate annual T-100 Domestic and International Freight Payload data by O-D market are also available at this site.

### **CFS Out-of-Scope Data**

“Auxiliary data” (discussed above) complement CFS Table 17 and allow missing cells within Table 17 to be estimated via Log-linear Modeling and Iterative Proportional Fitting (IPF). Waterborne Commerce, rail waybill, and Air Carrier data help to address some of the known weaknesses of the CFS survey in terms of mode coverage. The

Census Bureau “True Zero” data provide one approach to address Table 17 cells that are suppressed for disclosure and statistical reasons.

Other CFS gaps remain. Several commodities were totally absent in the 2002 CFS survey. In some cases, one or more shipments in a commodity’s supply chain were absent from the CFS survey. In other cases, whole categories of shipments were omitted from the survey, such as the movement of retail commodities from the point of final purchase to the home, business, etc. In yet other cases, there was evidence that the 2002 CFS undercounted some commodities and types of shipments – based on significant differences with other reliable data sources.

Earlier research suggested that previous CFS surveys undercounted total U.S. freight by a significant amount. A study by ORNL completed in 2000 estimated that the 1997 CFS captured only 75 percent of total U.S. freight shipments measured in tons, 74 percent when measured in ton-miles, and 81 percent when measured in value.<sup>1</sup>

As part of the 2002 FAF, a significant effort was launched to bridge the most serious of these CFS gaps. ORNL worked in collaboration with the Bureau of Transportation Statistics (BTS) and MacroSys Incorporated to estimate the following 15 CFS gaps and undercounts:

**Farm Based:** CFS omits shipments of farm commodities from the farm to the first point of sale, e.g. a grain elevator or a stockyard.

**Fisheries:** CFS omits shipments of fish and seafood from the boat at the dock to the processor or from the fish farm to the processor.

**Crude Petroleum:** Crude petroleum shipments are completely outside the scope of the 2002 CFS.

**Natural Gas:** Natural gas shipments are completely outside the scope of the 2002 CFS.

**Municipal Solid Waste (MSW):** MSW shipments are completely outside the scope of the 2002 CFS.

**Logging:** CFS omits shipments of logs from the point of harvest to the initial point of processing.

**Construction:** CFS does not cover shipments originating from the construction sector. The construction sector includes construction companies or establishments engaged in construction of residential and nonresidential buildings; utility systems; highway, street and bridge construction; and specialty trade contractors.

---

<sup>1</sup> Oak Ridge National Laboratory, (2000) *Freight USA. Highlights from the 1997 Commodity Flow Survey and Other Sources*. Report prepared for the Bureau of Transportation Statistics U.S. Department of Transportation, Washington, D.C 20590.

**Services:** The survey does not cover shipments originating from establishments involved in service industries. The missing services industries are finance and insurance; real estate, rental and leasing; professional, scientific and technical services; administrative and support, and waste management and remediation services; education services; health care and social assistance; arts, entertainment and recreation; accommodation and food services; other services (e.g., repair and maintenance, personal and laundry, religious, etc); and public administration. Also, the CFS does not include management of companies and enterprise services with the exception of corporate, subsidiary and regional managing offices.

**Publishing:** The CFS data gap on the publishing industry is primarily due to the adoption of the North American Industry Classification System (NAICS) in the 2002 CFS for selection of business establishments. In the 1997 and 1993 CFS businesses were selected based on their descriptions in the Standard Industry Classification (SIC).

**Retail:** CFS does not cover shipments originating from retail trade stores, including motor vehicle and parts dealers, furniture and home furnishings stores, electronics and appliance stores, building material and garden equipment and supplies dealers, food and beverage stores, health and personal care stores, gasoline stations, clothing and clothing accessories stores, sporting goods, book and music stores, general merchandise stores, florists, used merchandise, manufactured home dealers, etc.

**Household and Business Moves:** CFS does not capture freight movements by carriers that transport household and business furniture, equipment, etc.

**Imports:** Imports are completely outside the scope of the 2002 CFS. However, once import commodities enter the United States and change ownership, further shipments of those “imports” are captured within the CFS.

**Petroleum Products:** Petroleum products are technically within the scope of the CFS. However, previous research suggested that the 2002 CFS, like earlier editions, undercounted petroleum products.

**Exports:** The 2002 CFS collected data from U.S. business establishments located in the United States; thus the survey included exports from the United States by all freight modes. However, analysis of the 1993 and 1997 CFS export data suggests that the CFS underestimated U.S. export shipments.

**In-transits:** The CFS does not include shipments of commodities that originate outside of the United States, enter the United States by whatever mode, and then are shipped to some other country. Such shipments are called In-transits.

The 2002 CFS is estimated to have captured only 54 percent when measured in tons, 67 percent in ton-miles, and 63 percent in value.

# Methodology Overview

The methodology used to construct the 2002 FAF Commodity Origin-Destination Database includes several major steps and numerous assumptions. This section provides an overview of these steps and assumptions. The reader is referred to Report No. 3 for further details.

## **Step 1: CFS Table 17**

The starting point for the 2002 FAF is CFS Table 17. Note that Table 17 reports shipments at the level of 2-digit SCTG and the required mode detail. However, Table 17 reports shipments at the state level – not the FAF regional level. Our initial disaggregation from the state level to the FAF level was simple – i.e. divide shipments equally across all FAF regions that comprise each state. This simplistic assumption is followed up by adjustments per log-linear modeling and IPF as discussed below.

## **Step 2: Identify “True Zeros” in Table 17**

The Census Bureau identified those cells within Table 17 (disaggregated to the FAF regional level) for which the 2002 CFS had no samples. By assumption, those “true zero” cells were constrained to be “0”.

## **Step 3: Auxiliary Data and Conversions to SCTG**

Auxiliary data were obtained from USACE’s Waterborne Commerce database, Waybill data, and OAI’s air freight database. These data were converted to SCTG from the base commodity categorizations used by each data source. This allows a comparison of identical cells between Table 17 and non-CFS data for water, rail, and air freight. In those cases where direct comparisons were not possible at the four-dimensional level, comparisons were often possible at 2 or 3 dimensional levels. See Appendices #3, 4, and 5 for crosswalk details.

## **Step 4: Verify “True Zeros” with Auxiliary Data**

For those cells within Table 17 that were marked true zeros per Step 2, data constructed in Step 3 were compared to verify agreement between the two data sources. Waybill, Waterborne Commerce, and air freight data were compared for those particular cells to verify that neither of those datasets contradicts the true-zero conclusions based on Step 2. In cases of contradiction, i.e. where observations are found for cells previously marked as “true zero,” the restriction on that cell or margin was lifted.

## **Step 5: Augment Table 17 with Auxiliary Data**

Table 17, as modified per Steps 2 through 4, was augmented with water, rail, and air freight data as constructed in Step 3. In the cases of water, rail, and air, we have cells at

the 2, 3, and 4 dimensional levels from both CFS and from our auxiliary sources. Our Table 17 per Step 4 was augmented (e.g. skirted) with cells at the 2, 3, and 4 dimensional levels based on available auxiliary data. In other words, for some cells there are two values – one from the CFS and one from an auxiliary source. (Note that the auxiliary data were not included in marginal totals. Thus, marginal totals are constrained to be those from the CFS. Auxiliary data contribute to the log-linear step, but not to the IPF step.)

### **Step 6: Log-linear Modeling**

For those cells that have no observed value in CFS, as augmented with Census “true-zero” information, statistical procedures are available to estimate the most likely value of those missing cells, based upon statistical relationships extracted from cells with known values. For example, although CFS info is not available for fertilizer shipments from Iowa to Memphis, CFS information is available on the total fertilizer shipments from Iowa to all other FAF regions and for all commodities shipments from Iowa to Memphis. By examining the statistical relationships at higher orders of aggregation, a maximum likelihood value can be estimated for each missing cell. Please see Report 3 for further details about the exact formulation of the log-linear model.

Log-linear modeling was used to estimate these statistical relationships among FAF regions, modes, and commodities at 2, 3, and 4 dimensional levels. Log-linear models are specialized cases of general linear models. More specifically, log-linear analysis is an extension of the more familiar two-way contingency table in which the conditional relationship between two or more discrete variables is analyzed by taking the natural logarithm of the cell frequencies within the table. Log-linear models are a convenient way to analyze multi-dimensional contingency tables and estimate maximum likelihood values for missing cells.

In this study, the relationships among our discrete variables were based on CFS Table 17 with the following exceptions. In those cases where a relationship could not be calculated from Table 17, and for which a relationship could be calculated from auxiliary data, the auxiliary-based relationship was used in the log-linear model.

### **Step 7: Iterative Proportional Fitting**

Step 6 provides a complete four-dimensional matrix, but not a matrix that is consistent with totals at higher levels of aggregation. Iterative Proportional Fitting is a well accepted approach to adjust values within cells while maintaining the relationships between variables and assuring that rows and columns are consistent with the appropriate marginals. A “marginal” of a table is the set of quantities obtained by adding across all categories of any one or more of the cross classifying variables in a table.

The IPF procedure produces new estimates for each cell in the table at the 2, 3, and 4 dimensional levels such that they are in agreement with the marginal constraints, and is done so in an iterative fashion. In a two dimensional case, the elements of each row of



the table are prorated so that their totals equal the corresponding marginal; then the elements of each column are prorated so their totals equal their corresponding marginal. After this initial step, the estimates in the table no longer add across the rows to agree with the first marginal. The steps are repeated iteratively until the procedure converges to the unique solution that sums to the marginals while preserving the initial relationships between the variables in the table.

The product of Step 7 is a complete Table 17 (four dimensional) in which the initial values from Table 17 are maintained for those known cells at the 1, 2, 3, and 4 dimensional levels, including the true-zero values from Step 4.

(Note: The Log-linear and IPF steps actually involve seven or more dimensions – which calls for a more rigorous discussion. For example, the Log-linear step meshes tons, ton-miles, and value within one matrix and thus uses estimated statistical relationships across these three added dimensions to estimate missing cells. Please see Report 3 for details.)

### **Step 8: Adding Out-of-Scope Shipments**

The table derived in Step 7 does not include out-of-scope shipments. Those 15 categories of shipments, which were estimated in FAF<sup>2</sup> out-of-scope studies, must be added to the table from Step 7 to arrive at the final 2002 Commodity Origin-Destination Database. FAF out-of-scope flows were estimated initially at the national level in collaboration with BTS and MacroSys. These national totals were subsequently disaggregated by ORNL to the FAF regional level. Integrating these out-of-scope findings within our 2002 Commodity Origin-Destination Database is straightforward both conceptually and mechanically.

FAF<sup>2</sup> out-of-scope studies resulted in shipments at the appropriate four-dimensional level. Take MSW as an example. Shipments of MSW were estimated for all modes and all FAF regions – both intra and inter-regional shipments (e.g., from New York to the remainder of Pennsylvania – in terms of tons, ton miles, and value). MSW falls within SCTG category 41. Thus, the results of our out-of-scope studies are simply added to the appropriate cells to the tables resulting from Step 7.

### **Step 9: Analysis of Results**

The 2002 Commodity Origin-Destination Database contains 3 four-dimensional matrices (tons, ton miles, and value) for 43 commodities, 138 origins, 138 destinations, and 11 modes – for a total of more than 27 million cells. A multiyear effort is required to fully analyze these results, and is outside the scope of the current work. However, a first order analysis is required to assess reasonableness and to verify that the 2002 Commodity Origin-Destination Database is suitable for the next FAF step – i.e. network assignment.

## **Step 10: Validation**

Step 10 is validation. Three validation approaches are used. Cross validation is the first approach, in which random cells from the final 2002 Commodity Origin-Destination Database are removed and the 2002 Commodity Origin-Destination Database was re-estimated (Steps 1 through 8). The re-estimated tables are then compared to the tables from Step 8 using standard statistical approaches.

The second validation approach compares the statistical relationships among our parameters derived from Step 6 – i.e. from Table 17 as modified – with the same statistical relationships derived from our auxiliary data. The most significant statistical differences between the 2002 Commodity Origin-Destination Database and other auxiliary data sources can thus be identified and studied.

The third validation approach is similar to the second. Here we compare the absolute values of 2, 3, and 4 dimensional cells from Step 8 with known absolute values for those same cells from our auxiliary data sources.

## Appendix: Overview of the 2002 CFS

The following overview is taken from the BTS website:

[http://www.bts.gov/publications/commodity\\_flow\\_survey/2002/united\\_states\\_final/html/2002\\_commodity\\_flow\\_survey.html](http://www.bts.gov/publications/commodity_flow_survey/2002/united_states_final/html/2002_commodity_flow_survey.html)

### 2002 Commodity Flow Survey

#### GENERAL

The 2002 Commodity Flow Survey (CFS) is undertaken through a partnership between the U.S. Census Bureau, U.S. Department of Commerce, and the Bureau of Transportation Statistics (BTS), U.S. Department of Transportation. This survey produces data on the movement of goods in the United States. It provides information on commodities shipped, their value, weight, and mode of transportation, as well as the origin and destination of shipments of manufacturing, mining, wholesale, and select retail establishments. The data from the CFS are used by public policy analysts and for transportation planning and decision making to assess the demand for transportation facilities and services, energy use, and safety risk and environmental concerns. The CFS was last conducted in 1997.

This report contains background information on the 2002 Commodity Flow Survey and then presents detailed tabular results on shipment characteristics by mode of transportation, commodity, distance shipped, and shipment weight. In [Appendix A](#), key characteristics of the 2002 CFS are compared to those of the 1993 and 1997 surveys. [Appendix B](#) focuses on the reliability of the estimates and discusses sampling and nonsampling errors. Tables containing estimates of sampling variability corresponding to each table on shipment characteristics are also included in Appendix B.

This report presents the final United States summary data. It contains more detail than the preliminary United States report issued in December 2003 and reflects all revisions based on the geographic level analyses conducted since then. Additional reports will include data for census regions, divisions, states, and selected metropolitan areas, as well as selected data on exports and hazardous material shipments.

#### INDUSTRY COVERAGE

The 2002 CFS covers business establishments with paid employees that are located in the United States and are classified using the 1997 North American Industry Classification System (NAICS) in mining, manufacturing, wholesale trade, and select retail trade industries, namely, electronic shopping and mail-order houses. Establishments classified in services, transportation, construction, and most retail industries are excluded from the survey. Farms, fisheries, foreign establishments, and most government-owned establishments are also excluded.

The survey also covers auxiliary establishments (i.e., warehouses and managing offices) of multiestablishment companies, which have nonauxiliary establishments that are in-scope to the CFS or are classified in retail trade. The coverage of managing offices has been expanded in the 2002 CFS, compared to the 1997 CFS. For the 1997 CFS, the number of in-scope managing offices was reduced to a large extent based on the results of the 1992 Economic Census. A managing office was considered in-scope to the 1997 CFS only if it had sales or end-of-year inventories in the 1992 Census. However, research conducted prior to the 2002 CFS showed that not all managing offices with shipping activity in the 1997 CFS indicated sales or inventories in the 1997 Economic Census. Therefore, the 1997 Economic Census results were not used in the determination of scope for managing offices in the 2002 CFS.

For the 1993 CFS and the 1997 CFS, establishments were classified based on the 1987 Standard Industrial Classification System (SIC). Though an attempt was made to maintain similar coverage between the 1997 CFS and the

2002 CFS, there were some changes in industry coverage due to the conversion from SIC to NAICS. Most notably, coverage of the logging industry changed from an in-scope Manufacturing SIC code (SIC 2411) to an out-of-scope Agriculture, Forestry, Fishing, and Hunting NAICS code (NAICS 1133). Also, coverage of the publishing industry changed from in-scope Manufacturing SIC codes (SIC 2711, 2721, 2731, 2741, and part of 2771) to out-of-scope Information NAICS codes (NAICS 5111 and 51223).

## SHIPMENT COVERAGE

The CFS captures data on shipments originating from select types of business establishments located in the 50 states and the District of Columbia. The data do not cover shipments originating from business establishments located in Puerto Rico and other U.S. possessions and territories. Shipments traversing the U.S. from a foreign location to another foreign location (e.g., from Canada to Mexico) are not included, nor are shipments from a foreign location to a U.S. location. Imported products are included in the CFS at the point that they left the importer's domestic location for shipment to another location. Shipments that are shipped through a foreign territory with both the origin and destination in the U.S. are included in the CFS data. The mileages calculated for these shipments exclude the international segments (e.g., shipments from New York to Michigan through Canada do not include any mileages for Canada). Export shipments are included, with the domestic destination defined as the U.S. port, airport, or border crossing of exit from the U.S.

The "Industry Coverage" section of the text lists the NAICS groups covered by the CFS. Other industry areas that are not covered, but may have significant shipping activity, include agriculture and government. For agriculture, specifically, this means that the CFS does not cover shipments of agricultural products from the farm site to the processing centers or terminal elevators (most likely short-distance local movements), but does cover the shipments of these products from the initial processing centers or terminal elevators onward.

## MILEAGE CALCULATIONS

To estimate the distance traveled by each freight shipment sampled for the 2002 Commodity Flow Survey, the BTS Mileage Calculation Team used routing algorithms and an integrated, intermodal transportation network developed and updated expressly for this purpose by the Oak Ridge National Laboratory (ORNL). The BTS Team worked at a secure data site within the Census Bureau. Each record contained the ZIP Code shipment origin and destination, and the mode or modal sequence required by the routing algorithm for distance estimation. Each record also contained information on type of commodity moved, its weight, dollar value, and hazardous materials status. For export shipments, data on the U.S. port of exit were also identified, along with foreign destination city and country. Processing of shipment records began in the fall of 2002, with completion in October 2003.

One essential exercise was editing and imputing both absent and invalid geographic data elements, specifically origin and destination ZIP Codes, prior to estimating the distance traveled for each freight shipment. For this purpose, the BTS Mileage Calculation Team developed and maintained databases of domestic city/state names and foreign city/country names. The missing data elements, along with other related data problems found by the BTS Team, were either: (1) imputed because of high probability of accurate correction by the BTS Team, such as imputing a missing destination ZIP Code, given a destination city and state; or (2) reported back to the Census Bureau, allowing for call-backs to shippers for clarification/correction.

For a domestic shipment, the mileage is calculated between the center of the geographic area (centroid) of the U.S. origin ZIP Code and the centroid of the destination ZIP Code. The mileage for the shipments within a ZIP Code is calculated by means of a formula that approximates the longest distance within the boundaries of that ZIP Code. The mileage for an export shipment is calculated between a shipments centroid of U.S. origin ZIP Code and its foreign destination country (city in the case of Canada and Mexico), via a U.S. port of exit (POE), be it seaport, airport, or border crossing. However, only the portion of mileage that falls within the U.S. is included in the CFS estimates. That is to say, once the export reaches the POE, the POE is considered the final domestic destination, the domestic route is finished, and any following mileage is not counted from the POE. These mileages are computed using routing algorithms that find the minimum impedance path over mathematical representations of the U.S. and North American highway,

railway and waterway networks, and a transglobal representation of U.S. originating air freight and deep-sea transport networks. Shipment mileages were estimated for each record by summing over the distances of links contained within each minimum impedance path. Impedance was computed as a weighted combination of distance, time, and cost factors.

The ORNL multimodal network database is composed of mode-specific subnetworks representing each of the major transportation modes, such as highway, railway, waterway, and airway (pipeline network was not available due to security reasons). The links of these networks represent linehaul transportation facilities. Network nodes represent intersections and interchanges, along with the access points to the transportation network. To simulate local access, test links are created from each five-digit ZIP Code centroid to nearby nodes on the network. For the truck network, local access is assumed to exist everywhere. For the other modes this is not true. Before any test links are created for these modes, a search procedure is used to determine if and where such networks are most likely to provide access to the ZIP Code. For shipments involving more than one mode, such as truck-rail or rail-water shipments, intermodal transfer links are added to the network database to connect the individual modal networks together for routing purposes. An intermodal terminals database and a number of terminal transfer models were developed at ORNL to identify likely transfer points for different classes of freight. A measure of link impedance was calculated for each access, line-haul, and intermodal transfer link traversed by a shipment. These impedances were mode specific and are based on various link characteristics. For example, the set of links characterizing the highway network included speed impacting factors, such as the presence of a divided or undivided roadway, the degree of access control, the rural or urban setting, the number of lanes, the degree of urban congestion, and the length of the link. Link impedance measures were also assigned to the local access links. Intermodal transfer link impedances are estimated in terms of the time it takes to move goods through a transfer facility. In the case of rail and air freight, intercarrier transfer penalties were also considered to obtain proper route selections. A shortest path algorithm is used to find the minimum impedance path between a shipment's origin ZIP Code centroid and destination ZIP Code centroid. The cumulative length of the local access plus line-haul links on this path provides the estimated distances used in CFS mileage computations. When rail and air freight were involved, these shipment distances were often averaged over more than one path between an origin-destination pair.

## Mileage Data for Pipeline Shipments

For pipeline shipments, ton-miles and average miles per shipment are not shown in the tables. For most of these shipments, the respondents reported the shipment destination as a pipeline facility on the main pipeline network. Therefore, for the majority of these shipments, the resulting mileage represented only the access distance through feeder pipelines to the main pipeline network, and not the actual distance through the main pipeline network. Pipeline shipments are included in the U.S. totals for ton-miles and average miles per shipment. For security purposes, there is no pipeline network available in the public domain with which to route petroleum-based products. Hence, any modal distance, either single or multi, involving pipeline was considered as solely pipeline mileage from origin ZIP to destination ZIP and calculated to equal great circle distance (GCD). Note: Great circle distance is defined as the shortest distance between two points on the earth's surface, taking into account the earth's curvature.

## EXPLANATION OF TERMS

**Value of shipments.** The dollar value of the entire shipment. This was defined as the net selling value, f.o.b. plant, exclusive of freight charges and excise taxes. The value data are displayed in millions of dollars.

The total value of shipments, as measured by the CFS, and the U.S. gross domestic product (GDP) while similar in size provide different measures of economic activity in the United States and are not directly comparable. GDP is the value of all goods produced and services performed by labor and capital located in the United States. In 2002, the U.S. GDP was estimated at \$10.4 trillion (measured in current U.S. dollars). The value of shipments, as measured by the CFS, is the market value of goods shipped from manufacturing, mining, wholesale, and mail order retail establishments, as well as warehouses and managing offices of multiunit establishments.

Three important differences can be identified between GDP and value of shipments:

1. GDP captures goods produced by all establishments located in the United States, while the CFS measures goods shipped from a subset of all goods-producing establishments.
2. GDP measures the value of goods produced and of services performed. CFS measures the value of goods shipped.
3. GDP counts only the value-added at each step in the production of a product. CFS captures the value of shipments of materials used to produce or manufacture a product, as well as the value of shipments of the finished product itself. This means that the value of the materials used to produce a particular product contributes multiple times to the value.

**Commodity.** Products that an establishment produces, sells, or distributes. This does not include items that are considered as excess or byproducts of the establishment's operation. Respondents reported the description and the five-digit Standard Classification of Transported Goods (SCTG) code for the major commodity contained in the shipment, defined as the commodity with the greatest weight in the total shipment.

**Average miles per shipment.** For the 1993 CFS, we excluded shipments of Standard Transportation Commodity Classification (STCC) 27, Printed Matter, from our calculation of average miles per shipment. We made this decision after determining that respondents in the 1993 CFS shipping newspapers, magazines, catalogs, etc., had used widely varying definitions of the term "shipment." For the 1997 and 2002 CFS, we made numerous efforts throughout our data collection and editing to produce consistent results from establishments shipping SCTG 29, Printed Products. As a result, we have included printed products in the average miles per shipment estimates for 1997 and 2002.

**Distance shipped.** In Table 3, shipment data are presented for various "distance shipped" intervals. Shipments were categorized into these "distance shipped" intervals based on the great circle distance between their origin and destination ZIP Code centroids. All other distance-related data in this and other tables (i.e., ton-miles and average miles per shipment) are based on the mileage calculations. (See the "Mileage Calculations" section for more details.)

**Great circle distance.** The shortest distance between two points on the surface of a sphere over the surface of that sphere.

**Mode of transportation.** The type of transportation used for moving the shipment to its domestic destination. For exports, the domestic destination was the port of exit.

## Mode Definitions

In the instructions to the respondent, we defined the possible modes as follows:

1. **Parcel delivery/courier/U.S. Postal Service.** Delivery services that carry letters, parcels, packages, and other small shipments that typically weigh less than 100 pounds. Includes bus parcel delivery service.
2. **Private truck.** Trucks operated by a temporary or permanent employee of an establishment or the buyer/receiver of the shipment.
3. **For-hire truck.** Trucks that carry freight for a fee collected from the shipper, recipient of the shipment, or an arranger of the transportation.
4. **Railroad.** Any common carrier or private railroad.
5. **Shallow draft vessels.** Barges, ships, or ferries operating primarily on rivers and canals; in harbors, the Great Lakes, the Saint Lawrence Seaway; the Intra-coastal Waterway, the Inside Passage to Alaska, major bays and inlets; or in the ocean close to the shoreline.
6. **Deep draft vessel.** Barges, ships, or ferries operating primarily in the open ocean. Shipping on the Great Lakes and the Saint Lawrence Seaway is classified with shallow draft vessels.
7. **Pipeline.** Movements of oil, petroleum, gas, slurry, etc., through pipelines that extend to other establishments or locations beyond the shipper's establishment. Aqueducts for the movement of water are not included.
8. **Air.** Commercial or private aircraft, and all air service for shipments that typically weigh more than 100 pounds. Includes air freight and air express.
9. **Other mode.** Any mode not listed above.

10. **Unknown.** The shipment was not carried by a parcel delivery/courier/U.S. Postal Service, and the respondent could not determine what mode of transportation was used.

In the tables, we have used additional terms for mode, which we define as follows:

1. **Air (includes truck and air).** Shipments that used air or a combination of truck and air.
2. **Single modes.** Shipments using only one of the above-listed modes, except parcel or other and unknown.
3. **Multiple modes.** Shipments for which two or more of the following modes of transportation were used:
  - Private truck
  - For-hire truck
  - Rail
  - Shallow draft vessel
  - Deep draft vessel
  - Pipeline

In addition, Parcel, U.S. Postal Service, or Courier shipments are considered multiple modes because this category includes all parcel shipments whether on the ground or via air tendered to a parcel or express carrier. In defining this mode, we did not combine these shipments with any other reported mode because by their nature, Parcel, U.S. Postal Service or Courier are already multimodal. For example, if the respondent reported a shipment's mode of transportation as "parcel" and "air," we treated the shipment as parcel only. Also in the CFS reports, the "Truck and Rail" and "Rail and Water" combinations included under "Multiple Modes" may not reflect all the movement of trailers or containers by rail and at least one other mode of transportation. Since the shipper may not always know the modal combinations used to transport the goods, some shipments moving by more than one mode may be reported as a single mode shipment. This may result in underestimation of multimodal shipments in the CFS.

4. **Other multiple modes.** Shipments using any other mode combinations not specifically listed in the tables.
5. **Other and unknown modes.** Shipments for which modes were not reported, or were reported by the respondent as "Other" or "Unknown."
6. **Truck.** Shipments using for-hire truck only, private truck only, or a combination of for-hire truck and private truck.
7. **Water.** Shipments using shallow draft vessel only, deep draft vessel only, or Great Lakes vessel only. Combinations of these modes, such as shallow draft vessel and Great Lakes vessel are included as "Other multiple modes." (Note: By definition, "shallow draft," "Great Lakes," and "deep draft" are mutually exclusive.)
8. **Great Lakes.** In the tables in this publication, "Great Lakes" appears as a single mode. ORNL's transportation network and mileage calculation system allowed for separate mileage calculations for Great Lakes between the origin and destination ZIP Codes.

## Other Definitions and Terms

**Shipment.** A shipment is a single movement of goods, commodities, or products from an establishment to a single customer or to another establishment owned or operated by the same company as the originating establishment (e.g., a warehouse, distribution center, or retail or wholesale outlet). Full or partial truckloads are counted as a single shipment only if all commodities on the truck are destined for the same location. If a truck makes multiple deliveries on a route, the goods delivered at each stop are counted as one shipment. Interoffice memos, payroll checks, or business correspondence are not considered shipments. Shipments such as refuse, scrap paper, waste, or recyclable materials are not considered shipments unless the establishment is in the business of selling or providing these materials.

**Standard Classification of Transported Goods (SCTG).** The commodities shown in this report are classified using the SCTG coding system. The SCTG coding system was developed jointly by agencies of the United States and Canadian governments based on the Harmonized Commodity Description and Coding System (Harmonized System) to address statistical needs in regard to products transported. See [Appendix D](#) for more details.

**Ton-miles.** The shipment weight multiplied by the mileage traveled by the shipment. The respondents reported shipment weight in pounds. Aggregated pound-miles were converted to ton-miles. Mileage was calculated as the distance between the shipment origin and destination ZIP Codes. For shipments by truck, rail, or shallow draft vessels, the mileage excludes international segments. For example, mileages from Alaska to the continental United States exclude any mileages through Canada (see the “Mileage Calculations” section for more details). For trucks making multiple stops, the ton-miles are calculated for each delivery, and each drop-off point is treated as a final destination. Ton-miles estimates are displayed in millions.

**Tons shipped.** The total weight of the entire shipment. Respondents reported the weight in pounds. Aggregated pounds were converted to short-tons (2,000 pounds). For freight shipped to distribution centers for subsequent reshipment, the tonnage is counted each time the goods are transported.

**Total modal activity (Table 2 only).** The overall activity (e.g., ton-miles) of a specific mode of transportation, whether used in a single-mode shipment, or as part of a multiple-mode shipment. For example, the total modal activity for private truck is the total ton-miles carried by private truck in single-mode shipments, combined with the total ton-miles carried by private truck in all multiple-mode shipments that include private truck (private truck and for-hire truck, private truck and rail, private truck and air, etc.)